

## REMARKS

### *Claim Status*

Claims 1-35 were rejected. Claims 1-3, 5-9, 12-14, and 20-35 are amended. Support for the amendments can be found in the specification as originally submitted, particularly on pages 7-11. No claim is newly added or cancelled. No new matter is introduced. By this Amendment, claims 1-35 are pending.

### *Regarding Priority*

Applicants respectfully submit that, despite of informality, the provisional patent application 60/199,292 sufficiently discloses and adequately supports the foundational design of a new client/server architecture for text-to-speech (TTS) synthesis. These foundational technical features include:

- MIPS (million instructions per second) on client, memory on server;
- splitting speech data between client and server according to determinants;
- streaming in near real time speech data to the **client concatenator**; and
- utilizing the last in first out scheme so that minimal speech data need to be stored on client.

At the time of the invention, Matsumoto was the closest prior speech synthesis client/server system. However, as more specifically discussed and particularly distinguished in the present application, the speech quality of Matsumoto's system is limited because it utilizes formant synthesis and standard speech compression.

On the other hand, our new client/server architecture for TTS offers substantially high speech quality by splitting/dividing speech data, including corresponding processing and storage thereof, between client and server according to determinants such as frequency of usage. In other words, the server is designed to have loads of memory and to handle operations (i.e., processing steps) that would require large amounts of storage, while the client is designed to have high MIPS capability and to handle computationally intensive operations.

Any digital computer systems can be configured/programmed to implement the systems, methods and architecture disclosed therein. The techniques necessary to achieve this were well known to those skilled in the art. The independent claims of the present invention are amended herein to more prominently reflect the foundational features of the client/server architecture for TTS disclosed in and supported by the provisional patent application. Applicants therefore respectfully request that these foundational features recited in the claims be granted the priority date.

*Regarding Information Disclosure Statement*

Along with the application paperwork, applicants duly submitted a form PTO-1449, citing the reference listed at the end of page 3 (U.S. Pat. No. 5,940,796). A copy of the form PTO-1449, dated April 24, 2001, is attached herewith for the examiner's information. Applicants respectfully request that the examiner considers the reference listed thereon and returns a copy of the form after consideration.

*Regarding Specification*

Adopting the examiner's suggestions, the bottom paragraph of page 5 of the specification is amended herein to replace the term "simultaneously" with "concurrently" and to correct the misspelled word, "depends." The misspelled word, "concatenative," on page 3 is also corrected herein. Applicants thank the examiner for his thorough examination.

*Regarding Claim Objections*

Claims 5, 12, 15, 27, and 33 were objected to because it was not clear whether the term "compressed" recited therein pertains to storage, transmission or both. Claims 5, 12, 15, 27, and 33 are accordingly rephrased and/or amended herein to obviate the objections. No new matter is introduced.

In the present invention, the acoustic units *must* be compressed before transmission and *preferably* stored in such a compressed format. According the foundational design of the invention, the server has loads of memory, so the goal here is not necessarily to minimize storage space but to enhance performance by eliminating compression computations during

speech synthesis [Spec. page 4, line 31, to page 5, line 2; page 8, lines 27-33]. Alternatively, the acoustic units can also be stored as uncompressed acoustic units and only the selected acoustic units that correspond to the normalized text are compressed before transmission.

5 Accordingly, the preferred embodiment of the present invention has at least two advantages over prior systems – the server can transmit a lot more compressed speech data to the client and the database can store a lot more compressed speech data. None of the cited prior art teach or suggest an optimized compression method that offers these advantages.

10 Applicants further respectfully submit that the compression method taught and claimed in the present application is particularly *optimized to provide high compression while maintaining high quality of the transmitted acoustic units* [Spec. page 8, lines 32-33]. This is difficult to do because high compression (large amount of data reduction) usually means loss of speech quality, as discussed on page 3, lines 23-31.

15

Claim 15 was objected to because of an informality error. More particularly, the examiner contended that the verb “being” should have been “having been.” It is respectfully submitted that claim 15 recites three action steps handled by the client machine, i.e., a) receiving certain compressed acoustic units that correspond to a normalized text, b) decompressing the received acoustic units, and c) concatenating the decompressed acoustic units. The selection of the certain compressed acoustic units is implied but not explicitly recited because it is not part of the job of the client machine.

20

Thus, it is not entirely incorrect to use the verb “being” to describe what the compressed acoustic units *is*. Note the selection of compressed acoustic units has nothing to do with the compression method. Step a) of claim 15 recites, among others, what the client machine receives, i.e., compressed acoustic units that correspond to a normalized text, that are selected from a predetermined number of possible acoustic units, and that are compressed using a compression method. The compression method itself is optimized and selected in dependence on the predetermined number of possible acoustic units.

25

30

*Regarding Claim Rejections*

Claims 1-4, 6-11, 13-14, 23-25, 28-29, 30-32, and 34-35 were rejected under 35 U.S.C. § 103(a) as being unpatentable in view of Kochanski *et al.* (U.S. Pat. No. 6,625,576, hereinafter referred to as "Kochanski") in view of Matsumoto (U.S. Pat. 5,673,362). The  
5 rejections are respectfully traversed. Reconsideration is earnestly requested in view of the claim amendments presented herein and the following remarks.

Applicants respectfully submit that the present invention was conceived as far back as April 10, 2000. On April 24, 2000, applicants filed a provisional patent application sufficiently  
10 disclosing the foundational design of a new and inventive client/server architecture for text-to-speech synthesis. The invention was therefore constructively reduced to practice on April 24, 2000, rendering Kochanski inapplicable inasmuch as to the feature of "*in the client machine, concatenating the selected acoustic units.*"

As discussed in the background section of the specification, the client/server architecture in  
15 and by itself is not new. Matsumoto has already implemented a client/server TTS system. In April 2000, however, then existing TTS technology relied on powerful servers to perform computational intense speech signal synthesis. As exemplified in Matsumoto, the client receives and indiscriminately sends raw voice data to the voice synthesizing server. The  
20 voice synthesizing server processes the voice data, generates corresponding voice waveforms, and sends the voice waveforms back to the client. It was not known to split speech data, let alone splitting the corresponding processing steps thereof.

Moreover, Matsumoto does not teach or suggest concatenative synthesis. As discussed in the  
25 background section of the present application, concatenative synthesis was the more widely used high quality synthesis technique. However, because the quality of such a system is usually proportional to the size of the phonetic unit database and because it is less feasible for a single user to own and maintain such large databases, there was no incentive to perform concatenation at the client.

30 On the other hand, on or before April 24, 2000, applicants specifically designed a new

client/server architecture for TTS that, *inter alia*, purposely divides method steps required for high quality speech synthesis and designates a **client concatenator** for processing selected speech data *in the client*, predating Kochanski by more than nine months.

5 The disputable earlier date of invention notwithstanding, Kochanski still does not teach or suggest the invention as taught and claimed. As a whole, Kochanski *partitions* an otherwise **conventional** text-to-speech conversion algorithm into two portions: a “text analysis” portion executed exclusively on a server and a “speech synthesis” portion executed exclusively on a client. Kochanski does acknowledge that, to deliver a high quality text-to-speech system  
10 within the memory constraint of a client device, additional audio segments would need to be transmitted from the server [col. 10, lines 26-32]. However, Kochanski does not teach or suggest sending *highly compressed* acoustic units to the client machine.

At best, Kochanski teaches, in column 10, line 64 through column 11, line 2, that the server  
15 can transmit new segments *in a compressed form*, more specifically, “in the form of a reference to an existing cache item plus difference information,” *id.* Applicants respectfully submit that this is entirely different from and does not in any way suggest the particularly optimized compression method taught and claimed in the present application.

20 As taught and claimed in the present invention, the optimized compression method is based on the following framework: the complete set of possible acoustic units is known; each acoustic unit is divided into sequences of equal duration (chunks); each chunk is described by a set of parameters (e.g., line spectral pairs) according to a known model (e.g., linear predictive coding model). One of the parameters indicates the number of parameters used,  
25 e.g., the number of line spectral pairs used to describe a single chunk. The higher the number of parameters used, the more accurate will be the decompressed unit. The chunk is regenerated using the model and the parameter set, and a residual, the difference between the original and regenerated chunk, is obtained. The residual is modeled as, for example, a set of carefully placed impulses. The set of LPC parameters describing the full database can, in  
30 addition, be quantized into a small set of parameter vectors, or a codebook. The same

quantization can be performed on the residual vectors to reduce the description of each frame to two indices: that of the LPC vector and that of the residual vector.

Given this framework for the compression method, the method is optimized to select the number of parameters. Using a directed optimized search, the number of parameters for the frequency model and the number of impulse models for the residual are selected. The search is directed by an acoustic metric that measures quality. This metric is a combination of indirect measures such as a least mean squared difference between the encoded speech and the original, which can be used in, e.g., a gradient descent search, as well as perceptual measures such as a group of people grading the perceived quality, which post-qualifies a parameter set. The frequency model numbers and residual are then coded through an optimally selected codebook that uses the least possible number of code words to describe the known database. The indices to code words are the compressed acoustic units that are transmitted from the server to the client. This optimized compression method is our solution in providing high compression while maintaining high quality of the transmitted acoustic units.

It is respectfully submitted that one of ordinary skill in the art looking to the cited references would not be motivated and/or enabled to combine their teachings to achieve the claimed invention. What is lacking, given a fair reading of the references as a whole, is any teaching which would allow such a *high compression* of acoustic units while maintaining *high quality concatenation at the client*. Absent such teachings, it cannot be said that the claimed invention would have been obvious in view of the combined teachings of Kochanski and Matsumoto. Therefore, a *prima facie* case of obviousness has not been established.

### Conclusion

For the foregoing reasons, it is respectfully submitted that the present invention is patentably distinct from, not anticipated by, and unobvious in view of Kochanski and Matsumoto, individually and in combination. It is further respectfully submitted that, as amended, independent claims 1, 9, and 15 respectively recites subject matter not reached by the closest prior art of record under 35 USC § 103(a). Accordingly, independent claims 1, 9, and 15 are

submitted to be patentable. Reliance is placed on *In re Fine*, 5 USPQ 2d 1596, 1600 (Fed. Cir. 1988) and *Ex parte Kochan*, 131 USPQ 204 (Bd. App. 1960) for the allowance of dependent claims 2-8, 10-14, and 16-35, since they differ in scope from their respective parent independent claims 1, 9, and 15 which are submitted to be patentable.

5 This Response/Amendment is submitted to be complete and proper in that it places the present application in a condition for allowance without adding new matters. Since the examiner has done a thorough search in the first Office action in light of the entire application disclosure and claims, no new search would be necessary. Favorable  
10 consideration and a Notice of Allowance of all pending claims 1-35 are therefore earnestly solicited.

The examiner is sincerely invited to telephone the undersigned at 650-331-8413 for discussing an examiner's Amendment or any suggested actions for accelerating prosecution  
15 and moving the present application to allowance.

Respectfully submitted,



---

Katharina Wang Schuster, Reg. No. 50,000  
Attorney for the Applicants under 37 CFR 1.34

LUMEN INTELLECTUAL PROPERTY SERVICES  
2345 Yale Street, Second Floor  
Palo Alto, CA 94306  
(O) 650-424-0100 x 8413 (F) 650-424-0141